

О рандомизации статистик, приводящей к лучшей точности их приближений

Пучкин Никита Андреевич¹, Ульянов Владимир Васильевич²

¹ Национальный исследовательский университет “Высшая школа экономики”, Институт проблем передачи информации РАН, e-mail: npuchkin@hse.ru

² МГУ имени М.В.Ломоносова, e-mail: vulyanov@cs.msu.ru

Пусть имеются наблюдения $\mathbf{Y} = (Y_1, \dots, Y_r)$, представляющие из себя реализацию r -мерного случайного вектора с мультиномиальным распределением $\text{Mult}(n, \mathbf{p}_Y)$. Одной из фундаментальных задач математической статистики является проверка простой гипотезы $H_0 : \mathbf{p}_Y = \mathbf{p}$, где $\mathbf{p} = (p_1, \dots, p_r)$ – некоторый фиксированный вектор с неотрицательными компонентами, дающими в сумме 1. Предложен новый способ построения тестовой статистики, основанный на введении дополнительной рандомизации. Пусть $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ – случайный вектор с равномерным распределением на единичной сфере в \mathbb{R}^n . Представим вектор наблюдений в виде $\mathbf{Y} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 + \dots + \boldsymbol{\eta}_n$, где $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n$ – независимые одинаково распределенные случайные векторы с распределением $\text{Mult}(1, \mathbf{p}_Y)$. Вычислим $\mathbf{X}^\theta = \sum_{1 \leq i \leq n} \theta_i (\boldsymbol{\eta}_i - \mathbf{p})$. Предложенная статистика имеет вид

$$\mathcal{T}_\phi = \frac{2n}{\phi''(1)} \sum_{j=1}^r p_j \phi \left(1 + \frac{X_j^\theta}{\sqrt{np_j}} \right).$$

где $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ – выпуклая функция, удовлетворяющая условию

$$\phi(1) = \phi'(1) = 0, \quad \phi''(1) > 0.$$

Рандомизированная статистика \mathcal{T}_ϕ обладает лучшими асимптотическими свойствами по сравнению, например, с широко известными статистиками Пирсона [1] и Кресси-Рида [2]. В частности, доказано, что для любого $\delta \in (0, 1)$ с вероятностью не менее $(1 - \delta)$ расстояние Колмогорова между условным распределением $(\mathcal{T}_\phi | \boldsymbol{\theta})$ и $\chi^2(r - 1)$ убывает как $O((\log^4 n + \log^2(1/\delta))/n)$.

Работа Н. Пучкина выполнена при частичной финансовой поддержке конкурса “Молодая математика России”.

СПИСОК ЛИТЕРАТУРЫ

- [1] Pearson K. On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling // The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1900. Vol. 50, pp. 157–175.
- [2] Cressie N., Read T. Multinomial Goodness-of-Fit Tests // Journal of the Royal Statistical Society. Series B (Methodological). 1984. Vol. 46, pp. 440–464.